Can functional traits explain phylogenetic signal in the composition of a plant community?

Daijiang Li[1*], Anthony R. Ives[2], Donald M. Waller[1]

[1]Department of Botany, University of Wisconsin, 430 Lincoln Drive, Madison, Wisconsin, 53706

[2]Department of Zoology, University of Wisconsin, 430 Lincoln Drive, Madison, Wisconsin, 53706

Emails: daijianglee@gmail.com; arives@wisc.edu; dmwaller@wisc.edu

* Correspondence: Daijiang Li, Tel: (608) 265-2191

Authorship: DL and AI designed the study and performed the analyses. DL collected the vegetation and environmental data. DL and DW collected the functional trait data. DL wrote the first draft of the manuscript, and all authors worked together on revisions.

Running title: Functional traits and phylogeny

Key words: Functional traits, community assembly, phylogeny, phylogenetic linear mixed model.

Type of article: Letters

Words in abstract: 150; Words in main text: 4871

Number of references: 44; Number of figures: 2, Number of tables: 4

Abstract:

Phylogeny-based and functional trait-based analyses are used widely to study community composition. In principle, knowing all information about species traits should completely explain phylogenetic patterns in community composition. In reality, phylogenies may contain more information than the collection of measured traits. The extent to which functional trait information makes phylogenetic information redundant, however, is unknown. We used phylogenetic linear mixed models to analyze community composition of 55 understory plant species distributed across 30 forest sites in central Wisconsin. These communities showed strong phylogenetic attraction. Most of the 15 measured functional traits showed strong phylogenetic signal, but they only reduced the strength of phylogenetic community patterns in the abundances and presence/absences of co-occurring species by 57% and 89%, respectively, falling short of fully explaining phylogenetic community structure. Our study demonstrates the value of phylogenies in studying of community composition, especially with abundance data, even when rich functional trait data are available.

Introduction

Functional traits, arising as innovations through evolution, can capture essential aspects of species' morphology, ecophysiology, and life-history strategy (McGill *et al.* 2006; Violle *et al.* 2007). Although closely related species can differ greatly in some functional traits due to rapid evolution or ecological convergence (Losos, 2008, 2011), most functional traits show strong phylogenetic signal (Freckleton *et al.* 2002; Webb *et al.* 2002, Moles *et al.* 2005, Donoghue 2008). Functional traits, with or without phylogenetic signal, are known to influence the species composition of communities, thereby providing mechanistic links between fundamental

ecological processes and community structure (McGill *et al.* 2006; Violle *et al.* 2007; Adler *et al.* 2013). Functional traits also provide a common currency that facilitates comparisons among species and across regions, allowing us to assess the generality of patterns and predictions in community ecology (McGill *et al.* 2006). This has lead to a proliferation of studies using functional traits to understand community composition. Functional trait-based approaches, however, are limited by the fact that it is impossible to measure all potentially important functional traits affecting the distribution of species.

Even in the absence of functional trait information, it is still possible to infer the effects of (unmeasured) functional traits on community composition by investigating phylogenetic patterns in community composition. Phylogenies play an important role in community ecology by giving information about evolutionary relationships among species (Graves & Gotelli, 1993; Losos 1996; Baum & Smith, 2012). Because phylogenetically related species often share similar functional trait values, we expect phylogenetically related species to co-occur more often in the same communities reflecting their shared environmental tolerances. Conversely, if phylogenetically related species have similar traits that cause them to compete with each other, then closely related species may be less likely to co-occur. These and other processes relating functional traits to community composition likely lead to phylogenetic signatures in how species are distributed among communities (Webb *et al.* 2002). However, in principle, if we have information for all relevant functional traits, then we expect phylogeny to provide little additional information relevant for community composition. That is, when all of the functional traits affecting community composition are known, we do not expect the unexplained residual variation in the occurrence of species to have phylogenetic signal (Ives & Helmus, 2011).

In practice, we cannot obtain information about all relevant functional traits. In addition, phylogenetic signals in community composition may result from factors beyond functional traits, such as the biogeographical patterns generated as species disperse across a landscape (Ricklefs *et al.* 1993; Moen *et al.* 2009). If these forces are important, then even after accounting for all functional traits whose measurements are available, we should expect phylogenies to contain additional information about community composition (Vane-Wright *et al.* 1991; Cadotte *et al.* 2009). Thus far, however, we are aware of no study that has explicitly assessed the overlap between information from traits versus phylogeny. Here, we ask how much of the phylogenetic signal in the composition of a plant community assemblage can be explained by functional traits (Fig. 1).

We analyzed data on the abundance of 55 understory plant species distributed across 30 Wisconsin pine barrens sites (Li & Waller 2015). For each species, we had data on 15 functional traits and a recent highly resolved phylogeny (Cameron *et al. unpublished manuscript*[1]). At each site, we measured 20 environmental variables. Below, we first investigate whether there is phylogenetic pattern in community composition, using a phylogenetic community mixed model that tests for both "phylogenetic attraction" (phylogenetically related species more likely to occur in the same communities) and "phylogenetic repulsion." If there is phylogenetic pattern, then it could be produced by measured functional traits that themselves have phylogenetic signal (Fig. 1, arrows 2, 4, and 7), unmeasured functional traits with phylogenetic signal (Fig. 1, arrows 2, 5, and 8), or phylogenetic processes unrelated to functional traits (Fig. 1, arrow 6). We then developed a phylogenetic community mixed model incorporating the measured functional traits to ask whether there is phylogenetic signal in the residual variation in community composition

[1] Cameron, K., R. Kriebel, M. Pace, D. Spalink, P. Li, and K. Sytsma. *In prep*. A complete molecular community phylogeny for the flora of Wisconsin based on the universal plant DNA barcode.

after the effects of these traits are removed. This analysis tests the hypothesis that we can explain all of the phylogenetic pattern in community composition using measured functional traits. Finally, we use a phylogenetic community mixed model to investigate whether phylogenetically related species respond similarly to environmental gradients across the communities. The motivation for this final analysis is to indirectly identify possible unmeasured functional traits that might play a role in community assembly. In cases where phylogenetically related species respond similarly to an environmental gradient, species presumably share traits that confer similar tolerances to, or preferences for, specific environmental conditions. Thus, this final analysis could point towards additional functional traits that might be relevant for explaining patterns in community composition.

Methods

Data

Community composition. – We sampled 30 pine barrens forest sites in the central Wisconsin sand plains in 2012 using 50 1-m^2 quadrats placed along five transects at each site. Within each quadrat, we recorded the presence/absence of all understory vascular plant species (see Li & Waller 2015 for details). Across all sites, we recorded 152 species. For the analyses other than the initial exploration of phylogenetic patterns in community composition, we focused on the 55 species that occurred in three or more communities. We did this because we did not have functional trait data for many rare species, and we also wanted to limit the number of zeros in the data set.

Functional traits. – For the 55 focal species, we measured 11 continuous and four categorical functional traits on at least 12 individuals (four from each of at least three populations) using standard protocols (Pérez-Harguindeguy *et al.* 2013). Continuous traits include seed mass (*g*/seed), plant height (*cm*), specific leaf area (SLA, m^2/kg), leaf dry matter content (LDMC, %), leaf circularity (dimensionless), leaf length (*cm*), leaf width (*cm*), leaf thickness (*mm*), leaf carbon concentration (%), leaf nitrogen concentration (%), and stem dry matter content (SDMC, %). We aggregated categories of each categorical trait into two levels: growth form (woody vs. non-woody), life cycle (annual vs. non-annual), and pollination mode (biotic vs. abiotic). We divided seed dispersal mode into three binary variables (wind dispersed vs. not, animal dispersed vs. not, and unassisted vs. assisted dispersal). Collectively, these functional traits, covering the leaf-height-seed (LHS) plant ecology strategy (Westoby, 1998), represent multidimensional functions of plants associated with resource use, competitive ability, dispersal ability, etc. For analyses, we log-transformed highly skewed traits first and then Z-transformed the trait values to have means of zero and standard deviations of one, allowing coefficients in the mixed models to be interpreted as effect sizes.

Phylogeny. – The phylogeny used in this study is a subset of a phylogeny for all vascular plants in Wisconsin (Cameron *et al. unpublished manuscript*). Briefly, Cameron *et al.* used two plastid DNA barcode loci *rbcL* and *matK* to build the phylogeny using maximum likelihood (ML) in the program `RAXML` (Stamatakis, 2014). The phylogeny was then time-calibrated using the branch length adjuster (*bladj*) available in the program `phylocom` (Webb *et al.* 2008).

Environmental data. – At each site, we pooled six soil samples to measure the soil properties listed in Table 4. We also took six vertical fish-eye photographic images at each site to measure canopy cover. To characterize climatic conditions, we extracted daily precipitation and minimum

132 temperature for each site from interpolated values estimated by Kucharik *et al*. (2010) from 2002
133 to 2006 (data after 2006 were not available). All environmental variables were Z-transformed.

134 Phylogenetic community composition

135 We performed all analyses using both species abundances and species presence/absences among
136 communities. In the main text we present the analyses of abundance data, because including
137 abundance data in phylogenetic community analyses provides more information about
138 community assembly (Freilich & Connolly, 2015). In the Appendix we present the results for
139 presence/absence data.

140 We first tested for phylogenetic community structure without including environmental or
141 functional trait information. We used traditional metrics and randomization tests (i.e., null
142 models) to identify whether there was phylogenetic pattern (phylogenetic attraction or repulsion)
143 in the composition of our 30 communities. Specifically, we measured the phylogenetic structure
144 of species abundances at each site using phylogenetic species evenness (PSE, Helmus *et al.*
145 2007) and mean phylogenetic distance (MPD, Webb, 2000). For each site, we calculated PSE
146 and MPD, and then calculated the mean of these metrics ($\overline{\mathrm{PSE}}_{\mathrm{obs}}$ and $\overline{\mathrm{MPD}}_{\mathrm{obs}}$) across all 30
147 sites. To test for phylogenetic pattern, we permuted species randomly among sites (SIM2 in
148 Gotelli, 2000) 4999 times and then calculated metrics base on each permutation data set. If
149 $\overline{\mathrm{PSE}}_{\mathrm{obs}}$ or $\overline{\mathrm{MPD}}_{\mathrm{obs}}$ falls below (or above) 97.5% of the permutation values, then we infer a
150 statistically significant phylogenetic attraction (or repulsion). This null permutation model
151 retains the prevalence of each species across sites, but allows sites to change in species richness.
152 Using this null model where sites can vary in species richness is justified, because under the null
153 hypothesis of no phylogenetic signal, the values of PSE and MPD are independent of species

richness at the sites. We also performed permutation tests on the presence/absence of species from the 30 sites using phylogenetic species variation (PSV, Helmus *et al.* 2007) and MPD.

In addition to these permutation tests, we fit a phylogenetic linear mixed model (PLMM) to test for phylogenetic community patterns in species abundances. A PLMM establishes a flexible statistical base to subsequently incorporate functional trait and environmental variables. Furthermore, PLMMs tend to have greater statistical power than permutation tests (Ives & Helmus, 2011). To build the PLMM, let n be the number of species distributed among m sites. Letting Y be the $mn \times 1$ vector containing the abundance of species j ($j = 1, \ldots, n$) at site s ($s = 1, \ldots, m$), the PLMM is

$$\log(Y + 1) = \alpha + a_{\text{spp}[i]} + b_{\text{spp}[i]} + c_i + d_{\text{site}[i]} + e_i$$

$$a \sim \text{Gaussian}(\mathbf{0}, \sigma_a^2 \mathbf{I}_n)$$

$$b \sim \text{Gaussian}(\mathbf{0}, \sigma_b^2 \mathbf{\Sigma}_{\text{spp}})$$

$$c \sim \text{Gaussian}(\mathbf{0}, \text{kron}(\mathbf{I}_m, \sigma_c^2 \mathbf{\Sigma}_{\text{nested}}))$$

$$d \sim \text{Gaussian}(\mathbf{0}, \sigma_d^2 \mathbf{I}_m)$$

$$e \sim \text{Gaussian}(\mathbf{0}, \sigma_e^2 \mathbf{I}_{mn}) \qquad (1)$$

We use the convention of multilevel models here (Gelman & Hill, 2007), with fixed and random effects given by Greek and Latin letters, respectively. The function spp[i] maps the observation i in vector Y to the identity of the species (Gelman & Hill, 2007, p251-252), so i takes values from 1 to mn. The intercept α estimates the overall average log abundance of species across all sites. The following three random variables $a_{\text{spp}[i]}$, $b_{\text{spp}[i]}$ and c_i incorporate variation in abundance

among plant species. Specifically, the n values of $a_{spp[i]}$ give differences among species in mean log abundance across all sites and are assumed to be drawn independently from a Gaussian distribution with mean 0 and variance σ_a^2. The n values of $b_{spp[i]}$ also give differences in mean log abundance across sites but are assumed to be drawn from a multivariate Gaussian distribution with covariance matrix $\sigma_b^2\mathbf{\Sigma}_{spp}$, where the $n \times n$ matrix $\mathbf{\Sigma}_{spp}$ is derived from the phylogeny (see next paragraph), and the scalar σ_b^2 dictates the overall strength of the phylogenetic signal. Thus, $a_{spp[i]}$ and $b_{spp[i]}$ together capture variation in mean species log abundances that is either unrelated to phylogeny or has phylogenetic signal. The random variable c_i accounts for covariance in the log abundances of plant species nested within sites (using the Kronecker product, kron). Specifically, c_i assesses whether phylogenetically related plant species are more or less likely to co-occur at the same sites. Hence, c_i is used to measure either phylogenetic attraction or phylogenetic repulsion; because σ_c^2 dictates the overall strength of these phylogenetic patterns, it is the key term we are interested in. Random effect $d_{site[i]}$ is assumed to contain m values, one for each site, that are distributed by a Gaussian distribution with variance σ_d^2 to account for differences in the average log abundances of species from site to site. Finally, e_i captures residual variance σ_e^2.

We derived the phylogenetic covariance matrix $\mathbf{\Sigma}_{spp}$ from the assumption of Brownian motion evolution. If a continuous-valued trait evolves up a phylogenetic tree with a constant probability of slight increases or decreases, the covariance in trait values between two species will be proportional to the length of shared evolution, given by the distance on the phylogenetic tree between the root and the species' most recent common ancestor (Martins & Hanson 1997). This gives a direct way to convert the phylogeny into a hypothesis about the covariance matrix. For the assessment of phylogenetic attraction within sites, c_i, we use $\mathbf{\Sigma}_{nested} = \mathbf{\Sigma}_{spp}$. For phylogenetic

repulsion, we use the matrix inverse of $\mathbf{\Sigma}_{spp}$, $\mathbf{\Sigma}_{nested} = (\mathbf{\Sigma}_{spp})^{-1}$. Theoretical justification for $\mathbf{\Sigma}_{nested} = (\mathbf{\Sigma}_{spp})^{-1}$ comes from a model of competition among community members (Ives & Helmus 2011, Appendix A). Briefly, if the strength of competition between species is given by $\mathbf{\Sigma}_{spp}$, as might be the case if closely related species are more likely to share common resources, then the relative abundances of species will have covariance matrix $(\mathbf{\Sigma}_{spp})^{-1}$.

Equation 1 is the same as model I in Ives & Helmus (2011), except model I includes variation among species in mean log abundance across sites as fixed effects rather than two random effects, $a_{spp[i]}$ and $b_{spp[i]}$. This change allows us to align equation 1 with equation 3 (below) that includes variation in the relationship between trait values and log abundance within sites as random effects. In our analyses, treating variation among species in mean log abundance as fixed effects (results not presented) led to almost identical estimates of phylogenetic signal (estimates of σ_c^2), and therefore our treatment of $a_{spp[i]}$ and $b_{spp[i]}$ as random effects does not change the conclusions.

We fit the PLMM with maximum likelihood using function `communityPGLMM` in the `pez` (Pearse *et al.*, 2015) package of `R` (R Core Team, 2015). Statistical significance of the variance estimates σ^2 was determined using a likelihood ratio test. Because the null hypothesis $\sigma^2 = 0$ is on the boundary of the parameter space (σ^2 cannot be negative), we used the $0.5\chi_0^2 + 0.5\chi_1^2$ mixture distribution of Self & Liang (1987) for significance tests. The distribution of χ_0^2 represents a distribution with a point mass at 0, and the *p*-values given by the constrained likelihood ratio test are one-half the values that would be calculated from a standard likelihood ratio test using χ_1^2. Simulations suggest that *p*-values calculated in this way are more conservative (have higher values) than those from a parametric bootstrap (Appendix Text S1).

Our data set contained many zeros (Fig. 2), raising the question of the validity of applying a linear model to transformed data. Nonetheless, transforming data and applying a linear analysis is robust when assessing the significance of regression parameters (Ives, 2015).

Can functional traits explain phylogenetic community composition?

To quantify how much of the variation in phylogenetic patterns can be explained by measured functional traits, we estimated PLMMs with and without functional traits, and then compared the strength of phylogenetic signal in the residual variation: if functional traits alone serve to explain phylogenetic community composition, then as functional traits are included, the strength of the phylogenetic signal in the residuals should decrease. We selected functional traits one by one based on the two conditions necessary for them to generate phylogenetic signal in community composition. First, a functional trait must show phylogenetic signal among species, because in the absence of phylogenetic signal among species, a trait could not produce phylogenetic signal in species' abundances. Second, there must be variation among sites in the relationship between species trait values and abundances; if a trait has phylogenetic signal but there is no variation in relationships between plant functional trait values and abundances among sites, then it will contribute to the overall phylogenetic signal of species abundance and will be captured by $b_{\text{spp}[i]}$ in equation 1, but it will not affect phylogenetic co-occurrence patterns captured by c_i. Therefore, we only investigate traits that exhibit both strong phylogenetic signal and variation among sites in the apparent advantages the traits give to species.

We tested the phylogenetic signal for each functional trait using model-based methods. Each continuous trait was tested with Pagel's λ (Pagel, 1999) using `phylolm` (Ho & Ané, 2014). For the binary traits, we applied phylogenetic logistic regression (Ives & Garland, 2010) as

implemented by `phyloglm` (Ho & Ané, 2014). We also tested phylogenetic signal of functional traits via Blomberg's *K* (Blomberg *et al*. 2003) with `picante` (Kembel *et al*. 2010).

We tested variation of relationships between trait values and log abundances with the LMM

$$\log(Y + 1) = \alpha + a_{\text{spp}[i]} + (\beta + b_{\text{site}[i]})t_{\text{spp}[i]} + e_i$$

$$a \sim \text{Gaussian}(\mathbf{0}, \sigma_a^2\mathbf{I}_n)$$

$$b \sim \text{Gaussian}(\mathbf{0}, \sigma_b^2\mathbf{I}_m)$$

$$e \sim \text{Gaussian}(\mathbf{0}, \sigma_e^2\mathbf{I}_{mn}) \quad (2)$$

where $t_{\text{spp}[i]}$ is the focal functional trait value of the species corresponding to observation i, and σ_b^2 gives the variation among sites in the relationship between species trait values and log abundances. This formulation is closely related to the model used by Pollock *et al.* (2012). If $\sigma_b^2 > 0$, we conclude that different sites select species differently based on the tested trait. We use $p < 0.1$ here to lower the risk of excluding potential important functional traits.

We quantified the contribution of a trait to the observed phylogenetic pattern in community composition using the model

$$\log(Y + 1) = \alpha + a_{\text{spp}[i]} + b_{\text{spp}[i]} + c_i + d_{\text{site}[i]} + (\beta + f_{\text{site}[i]})t_{\text{spp}[i]} + e_i$$

$$a \sim \text{Gaussian}(\mathbf{0}, \sigma_a^2\mathbf{I}_n)$$

$$b \sim \text{Gaussian}(\mathbf{0}, \sigma_b^2\mathbf{\Sigma}_{\text{spp}})$$

$$c \sim \text{Gaussian}(\mathbf{0}, \text{kron}(\mathbf{I}_m, \sigma_c^2\mathbf{\Sigma}_{\text{nested}}))$$

$$d \sim \text{Gaussian}(\mathbf{0}, \sigma_d^2\mathbf{I}_m)$$

$$f \sim \text{Gaussian}(\mathbf{0}, \sigma_f^2 \mathbf{I}_m)$$

$$e \sim \text{Gaussian}(\mathbf{0}, \sigma_e^2 \mathbf{I}_{mn}) \quad (3)$$

This model is the same as equation 1 used to assess phylogenetic patterns in community composition, except that it includes functional trait values $t_{spp[i]}$. The proportion of phylogenetic signal in species composition (estimated by σ_c^2) that trait $t_{spp[i]}$ can explain is assessed by comparing σ_c^2 between models with and without this trait as a product with the random effect $f_{site[i]}$. Finally, to evaluate the overall contribution of functional traits to the observed phylogenetic patterns, we built a multivariate version of equation 3 which included all traits that have both phylogenetic signal and strong variation among sites.

Does any environmental variable drive phylogenetic pattern?

If phylogenetic patterns in community composition are observed, yet no functional traits can explain the patterns, how could we identify additional functional traits that might be responsible? Phylogenetically related species usually are assumed to be ecologically similar due to niche conservatism (Wiens *et al.* 2010). Therefore, related species will tend to have similar responses to environmental variables. If these environmental variables are strong enough to drive phylogenetic patterns in community composition, then functional traits that are associated with tolerance or sensitivity to these environmental variables will likely be important in explaining community composition. Thus, we investigated phylogenetic patterns in the responses of species to environmental variables to suggest additional, unmeasured functional traits that might be important to explain phylogenetic patterns in community composition.

We tested for phylogenetic patterns in the responses of species to environmental variables using the PLMM

$$\log(Y + 1) = \alpha + a_{spp[i]} + b_{spp[i]} + (\beta + g_{spp[i]} + h_{spp[i]})x_{site[i]} + e_i$$

$$a \sim \text{Gaussian}(\mathbf{0}, \sigma_a^2\mathbf{I}_n)$$

$$b \sim \text{Gaussian}(\mathbf{0}, \sigma_b^2\mathbf{\Sigma}_{spp})$$

$$g \sim \text{Gaussian}(\mathbf{0}, \sigma_g^2\mathbf{I}_n)$$

$$h \sim \text{Gaussian}(\mathbf{0}, \sigma_h^2\mathbf{\Sigma}_{spp})$$

$$e \sim \text{Gaussian}(\mathbf{0}, \sigma_e^2\mathbf{I}_{mn}) \quad (4)$$

Here, $g_{spp[i]}$ and $h_{spp[i]}$ represent non-phylogenetic and phylogenetic variation among species in their response to environmental variable x (see model II in Ives & Helmus, 2011). The key parameter of interest is σ_h^2, which we tested using a likelihood ratio test. If $\sigma_h^2 > 0$, phylogenetically related species respond to environmental variable x in similar ways, suggesting the existence of an unmeasured phylogenetically inherited trait that is associated with species tolerances or sensitivities to x. Given the large number of environmental variables in our data set, we first applied equation 4 without the term $b_{spp[i]}$ and $h_{spp[i]}$, and selected environmental variables for which there was variation in responses among species given by $g_{spp[i]}$ regardless of whether this variation was phylogenetic. For variables x for which $\sigma_g^2 > 0$ in the reduced version of equation 4, we then applied the full equation 4 and tested whether $\sigma_h^2 > 0$.

Results

Phylogenetic community composition

Phylogenetically related species co-occurred more often than expected by chance in pine barrens communities in central Wisconsin (Fig. 2). Permutation tests including all 152 species showed that closely related species are likely to have positive covariances in abundance among communities, as judged by either phylogenetic species evenness ($\overline{\mathrm{PSE}}_{\mathrm{obs}} = 0.32$, $p = 0.03$) or mean phylogenetic distance ($\overline{\mathrm{MPD}}_{\mathrm{obs}} = 338$, $p = 0.01$). In contrast, when we confine analyses to the 55 focal species that occurring in at least three communities, the permutation tests failed to show statistically significant phylogenetic patterns (abundance data: $\overline{\mathrm{PSE}}_{\mathrm{obs}} = 0.27$, $p = 0.29$; $\overline{\mathrm{MPD}}_{\mathrm{obs}} = 286$, $p = 0.17$; presence/absence data: $\overline{\mathrm{PSE}}_{\mathrm{obs}} = 0.31$, $p = 0.20$; $\overline{\mathrm{MPD}}_{\mathrm{obs}} = 342$, $p = 0.20$). Nevertheless, the PLMM ($p = 0.008$; Table 1) and PGLMM ($p < 0.001$; Appendix Table S1) both reveal statistically significant phylogenetic patterns for the 55 focal species.

Can functional traits explain phylogenetic community composition?

Most functional traits showed strong phylogenetic signal (Table 2). Five traits – leaf width, leaf thickness, SLA, leaf circularity, and animal dispersal (marginally significant) – also significantly affected plant species' abundances among sites ($\sigma_b^2 > 0$, equation 2, Table 2), indicating that different sites selected different species based on these three functional traits. Individually, the five traits reduced the phylogenetic variance in community composition (as measured by reduction in σ_c^2 in equation 3 when including these traits) by 18%, 8%, 7%, 2%, and 1%, respectively. Traits that did not pass our two-steps selection individually explained negligible amount of the phylogenetic variance (all <1% and mostly ~0%, data not shown), verifying our

initial selection of traits. Including all five traits in the final model reduces the phylogenetic variation σ_c^2 by 57%. Thus, the many functional traits we measured in this study can only reduce the phylogenetic signal in community composition by 57%. Converting the data to presence/absence and using the PGLMM equivalent of equation 3 reduces σ_c^2 by 89% (Appendix, Table S3). Thus, functional traits explained more of the phylogenetic patterns in the presence/absence of species from communities than in their log abundance, although functional traits still cannot fully explain the phylogenetic pattern in community composition.

Does any environmental variable drive phylogenetic pattern?

There was significant variation among species in their responses to most of the environmental variables we measured, including soil conditions, canopy shade, precipitation, and minimum temperature (Table 4). However, there was no phylogeny signal in the differences among species in their responses to these variables (last column in Table 4). Therefore, no environmental variables we measured can explain the observed phylogenetic pattern in community composition. Using the PGLMM with the presence/absence data, species' responses to minimum temperature and soil pH, Ca, and Mn concentration all show phylogenetic signal. That is, related species tend to occupy similar sites as measured by these environmental variables (Appendix Table S4). Therefore, functional traits associated with these environmental variables could potentially be responsible for phylogenetic patterns in presence/absence of species among communities.

Discussion

We used our extensive database of functional traits to answer a key question in trait-based and phylogeny-based community ecology: Can information about functional traits explain

phylogenetic patterns in community composition? Phylogenetically related plant species are more likely to reach similar abundances in the same pine barren communities of central Wisconsin, yet we could not explain this pattern completely using information about species' functional traits. When functional traits that themselves showed phylogenetic signal among species were included in the phylogenetic linear mixed model (PLMM) for log abundances of species in communities, that component of the residual variance having phylogenetic covariances decreased by only 57%. The decrease in the phylogenetic component of residual variation was 89% in the analyses of presence/absence data, yet even this leaves residual phylogenetic pattern in the unexplained variation in the presence/absence of species among communities. Thus, even though we measured 15 functional traits, including most of the standard functional traits used to analyze plant community structure, we could not fully explain the phylogenetic patterns in community composition. This suggests that there are either important functional traits that we have not measured, or that there are phylogenetic processes unrelated to functional traits that we have not identified. In either case, these results suggest that including phylogenetic information in addition to functional traits provides further insights into the processes affecting community assembly.

When using the subset of 55 species that occurred in three or more communities, the PLMM (and PGLMM), but not permutation tests, found statistically significant phylogenetic patterns. Ives & Helmus (2011) showed that phylogenetic mixed models have greater statistical power than the metrics like PSE and MPD used with permutation tests. Simulations (Appendix Text S1) show that PLMM analyses tended to have, if anything, incorrectly low Type I error rates, implying that our PLMM results were not the result of false positives. We can thus conclude that

363 closely related species are more likely to co-occur and share similar abundances than expected
364 by chance in these pine barren communities.

365 Incorporating functional traits reduced the phylogenetic component of residual variation in
366 species composition, what could explain the remaining phylogenetic component? Some
367 unknown historical process might account for this residual phylogenetic variation (Fig. 1B, IV).
368 However, our sites are all located within 100 km with each other, making it unlikely that
369 historical biogeographical processes strongly affect the composition of these communities. It
370 seems more likely that the main source of phylogenetic patterns that were not explained by our
371 measured functional traits is additional unmeasured functional traits. Further analyses of the
372 presence/absence data using PGLMMs suggested that soil conditions (pH, Ca, and Mn levels)
373 and climate (minimum temperature) are potential driving variables for the residual phylogenetic
374 patterns (Appendix Table S3). Traits associated with plant responses to these gradients in
375 environmental conditions could thus account for more of the residual phylogenetic patterns. The
376 functional traits we measured, however, are traits that are unlikely to capture species-specific
377 responses to soil and climatic conditions, and we do not have information on likely traits such as
378 root structure, micorrhizal associations, frost tolerance, etc. We expect such traits might be able
379 to explain more of the phylogenetic pattern in community composition.

380 We found that functional traits could explain a greater part of the phylogenetic component of the
381 pattern of species presence/absence (89%) than of species abundances (57%). This is unlikely to
382 be a statistical artifact. Because we used only the most common 55 species, detection of species
383 in sites where they occur is likely to be high. In contrast, we expect considerable within-species
384 variation in our estimates of abundance. Because within-species variation will decrease
385 phylogenetic signal (Ives *et al.* 2007), we would expect less residual phylogenetic variation in

386 the abundance data than in the presence/absence data, the opposite of what we found. Therefore,
387 our results suggest that the functional traits we measured have a greater effect on the overall
388 suitability of sites for species than the finer-tuned quality of the sites to support large
389 populations, supporting the argument that including abundance data in phylogenetic community
390 analyses provides more information about community assembly (Freilich & Connolly, 2015).

391 ## Implications

392 Our results have several implications for community ecology. First, it is clear that studying
393 community composition should incorporate analyses of both phylogenetic structure and
394 functional traits. Phylogenetic and trait information clearly complement each other in allowing
395 sophisticated analyses that can partition the amount of phylogenetic signal in community
396 composition that is associated with functional trait variation (Fig. 1). Our results provide
397 empirical support from community ecology for the argument that phylogenies can provide more
398 information than a set of discretely measured traits (Vane-Wright *et al.* 1991; Cadotte *et al.*
399 2009). Although functional traits are necessary to accurately infer the processes driving
400 phylogenetic patterns (Kraft *et al.* 2007; Cavender-Bares *et al.* 2009), functional traits alone may
401 often fail to provide a complete picture of community structure.

402 Second, model-based methods are being increasingly applied in ecology because they are more
403 interpretable, flexible, and powerful than either null models or conventional algorithmic
404 multivariate analyses (Warton *et al.* 2014). With phylogenetic linear mixed models (PLMM), we
405 not only detected phylogenetic patterns in community composition, but also assessed the extent
406 to which these could be explained by functional traits. The ability to combine both phylogenies
407 and functional traits into the same statistical model using PLMMs (and PGLMMs) provides an

integrated and quantitative framework for analyzing ecological communities and predicting abundance of one taxon from others.

Finally, we can use phylogenetic analyses to suggest possible unmeasured functional traits that underlie patterns in community composition and that therefore should be measured. If species respond differently to an environmental variable, and if these differences are phylogenetic (i.e., related species respond to the environmental variable in similar ways), then there is likely to be a functional trait or traits that underlie the response of species to this environmental variable. In our study, the phylogenetic patterns in species responses to edaphic conditions like soil chemistry highlighted our lack of data on the specific functional traits related to roots or water/nutrient uptake. While this reveals that our study is incomplete, it also provides a valuable lesson and demonstrates the power of the integrated PLMM approach.

Acknowledgements

We thank K. Cameron, R. Kriebel, M. Pace, D. Spalink, P. Li, and K. Sytsma for building and providing the phylogeny we used in this study. This project was funded by US-NSF grant DEB-1046355 and DEB-1240804.

References:

Adler, P.B., Fajardo, A., Kleinhesselink, A.R. & Kraft, N.J.B. (2013). Trait-based tests of coexistence mechanisms. *Ecol. Lett.*, 16, 1294–1306.

Baum, D.A. & Smith, S.D. (2012). *Tree Thinking: An Introduction to Phylogenetic Biology*. 1st Edition. Roberts Company Publishers, Greenwood Village, Colo.

Blomberg, S.P., Garland, T. & Ives, A.R. (2003). Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution*, 57, 717–745.

Cadotte, M.W., Cavender-Bares, J., Tilman, D. & Oakley, T.H. (2009). Using Phylogenetic, Functional and Trait Diversity to Understand Patterns of Plant Community Productivity. *PLoS ONE*, 4, e5695.

Cavender-Bares, J., Kozak, K.H., Fine, P.V.A. & Kembel, S.W. (2009). The merging of community ecology and phylogenetic biology. *Ecol. Lett.*, 12, 693–715.

Donoghue, M.J. (2008). A phylogenetic perspective on the distribution of plant diversity. *Proc. Natl. Acad. Sci. U.S.A.*, 105, 11549–11555.

Freckleton, R.P., Harvey, P.H. & Pagel, M. (2002). Phylogenetic Analysis and Comparative Data: A Test and Review of Evidence. *Am. Nat.*, 160, 712–726.

Freilich, M.A. & Connolly, S.R. (2015). Phylogenetic community structure when competition and environmental filtering determine abundances. *Global Ecol. Biogeogr.*, 24, 1390–1400.

Gelman, A. & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Gotelli, N.J. (2000). Null model analysis of species co-occurrence patterns. *Ecology*, 81, 2606–2621.

Graves, G.R. & Gotelli, N.J. (1993). Assembly of avian mixed-species flocks in Amazonia. *Proc. Natl. Acad. Sci. U.S.A.*, 90, 1388–1391.

Helmus, M.R., Bland, T.J., Williams, C.K. & Ives, A.R. (2007). Phylogenetic Measures of Biodiversity. *Am. Nat.*, 169, E68–E83.

Ho, L.S.T. & Ané, C. (2014). A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Syst. Biol.*, syu005.

Ives, A.R. (2015). For testing the significance of regression coefficients, go ahead and log-transform count data. *Methods Ecol. Evol.*, 6, 828–835.

Ives, A.R. & Garland, T. (2010). Phylogenetic Logistic Regression for Binary Dependent Variables. *Syst. Biol.*, 59, 9–26.

Ives, A.R. & Helmus, M.R. (2011). Generalized linear mixed models for phylogenetic analyses of community structure. *Ecol. Monogr.*, 81, 511–525.

Ives, A.R., Midford, P.E. & Garland, T. (2007). Within-Species Variation and Measurement Error in Phylogenetic Comparative Methods. *Syst. Biol.*, 56, 252–270.

460 Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H. & Ackerly, D.D.*et al.*
461 (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, 26, 1463–1464.

462 Kraft, N.J.B., Cornwell, W.K., Webb, C.O. & Ackerly, D.D. (2007). Trait Evolution,
463 Community Assembly, and the Phylogenetic Structure of Ecological Communities. *Am. Nat.*,
464 170, 271–283.

465 Kucharik, C.J., Serbin, S.P., Vavrus, S., Hopkins, E.J. & Motew, M.M. (2010). Patterns of
466 Climate Change Across Wisconsin From 1950 to 2006. *Phys. Geogr.*, 31, 1–28.

467 Li, D. & Waller, D. (2015). Drivers of observed biotic homogenization in pine barrens of central
468 Wisconsin. *Ecology*, 96, 1030–1041.

469 Losos, J.B. (1996). Phylogenetic Perspectives on Community Ecology. *Ecology*, 77, 1344–1354.

470 Losos, J.B. (2008). Phylogenetic niche conservatism, phylogenetic signal and the relationship
471 between phylogenetic relatedness and ecological similarity among species. *Ecol. Lett.*, 11, 995–
472 1003.

473 Losos, J.B. (2011). Seeing the Forest for the Trees: The Limitations of Phylogenies in
474 Comparative Biology. *Am. Nat.*, 177, 709–727.

475 Martins, E.P. & Hansen, T.F. (1997). Phylogenies and the Comparative Method: A General
476 Approach to Incorporating Phylogenetic Information into the Analysis of Interspecific Data. *Am.*
477 *Nat.*, 149, 646–667.

478 McGill, B.J., Enquist, B.J., Weiher, E. & Westoby, M. (2006). Rebuilding community ecology
479 from functional traits. *Trends Ecol. Evol.*, 21, 178–185.

480 Moen, D.S., Smith, S.A. & Wiens, J.J. (2009). Community Assembly Through Evolutionary
481 Diversification and Dispersal in Middle American Treefrogs. *Evolution*, 63, 3228–3247.

482 Moles, A.T., Ackerly, D.D., Webb, C.O., Tweddle, J.C., Dickie, J.B. & Westoby, M. (2005). A
483 Brief History of Seed Size. *Science (80-)*, 307, 576–580.

484 Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, 401, 877–884.

485 Pearse, W.D., Cadotte, M.W., Cavender-Bares, J., Ives, A.R., Tucker, C.M. & Walker, S.C.*et al.*
486 (2015). Pez: Phylogenetics for the environmental sciences. *Bioinformatics*, 31, 2888–2890.

487 Pérez-Harguindeguy, N., Díaz, S., Garnier, E., Lavorel, S., Poorter, H. & Jaureguiberry, P.*et al.*
488 (2013). New handbook for standardised measurement of plant functional traits worldwide. *Aust.*
489 *J. Bot.*, 61, 167–234.

490 Pollock, L.J., Morris, W.K. & Vesk, P.A. (2012). The role of functional traits in species
491 distributions revealed through a hierarchical model. *Ecography*, 35, 716–725.

R Core Team. (2015). R: A Language and Environment for Statistical Computing.

Ricklefs, R.E. & Schluter, D. (1993). Species diversity in ecological communities: Historical and geographical perspectives. In: *Species diversity in ecological communities: Historical and geographical perspectives.* University of Chicago Press.

Self, S.G. & Liang, K.-Y. (1987). Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions. *J. Am. Stat. Assoc.*, 82, 605–610.

Stamatakis, A. (2014). RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics*, btu033.

Vane-Wright, R.I., Humphries, C.J. & Williams, P.H. (1991). What to protect? Systematics and the agony of choice. *Biol. Conserv.*, 55, 235–254.

Violle, C., Navas, M.L., Vile, D., Kazakou, E., Fortunel, C. & Hummel, I.*et al.* (2007). Let the concept of trait be functional! *Oikos*, 116, 882–892.

Warton, D.I., Foster, S.D., Death, G., Stoklosa, J. & Dunstan, P.K. (2014). Model-based thinking for community ecology. *Plant Ecol.*, 1–14.

Webb, C.O. (2000). Exploring the phylogenetic structure of ecological communities: An example for rain forest trees. *Am. Nat.*, 156, 145–155.

Webb, C.O., Ackerly, D.D. & Kembel, S.W. (2008). Phylocom: Software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics*, 24, 2098–2100.

Webb, C.O., Ackerly, D.D., McPeek, M.A. & Donoghue, M.J. (2002). Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.*, 33, 475–505.

Westoby, M. (1998). A leaf-height-seed (LHS) plant ecology strategy scheme. *Plant Soil*, 199, 213–227.

Wiens, J.J., Ackerly, D.D., Allen, A.P., Anacker, B.L., Buckley, L.B. & Cornell, H.V.*et al.* (2010). Niche conservatism as an emerging principle in ecology and conservation biology. *Ecol. Lett.*, 13, 1310–1324.

Tables:

Table 1 Estimated variance of random effects for the PLMM (equation 1) used to detect phylogenetic patterns in community composition.

PLMM	σ^2_a	σ^2_b	σ^2_c	σ^2_d	σ^2_e	$p(\sigma^2_c = 0)$	AIC
Phylogenetic attraction: $c \sim \text{Gaussian}(\mathbf{0}, \text{kron}(\mathbf{I}_m, \sigma^2_c\mathbf{\Sigma}_{spp}))$	0.98	0	6.50×10^{-3}	0	0.5154	0.008	3900
Phylogenetic repulsion: $c \sim \text{Gaussian}(\mathbf{0}, \text{kron}(\mathbf{I}_m, \sigma^2_c\,(\mathbf{\Sigma}_{spp})^{-1})$	0.98	0	0	2.28×10^{-2}	0.5308	0.496	3906
Non-nested model: c removed	0.98	2.29×10^{-2}	-	0	0.5306	-	3904

527 **Table 2** Phylogenetic signal and site variation for each functional trait. P-values for the null hypothesis σ_b^2
528 = 0 (equation 2) implying no difference among sites in the effects of trait values on log abundance are
529 given in the column labeled $p(\sigma_b^2 = 0)$. Functional traits with strong phylogenetic signal and $p(\sigma_b^2 = 0)$ <
530 0.1 are considered to be important in explaining phylogenetic patterns.

Trait	Pagel's λ	*K*	$p(\sigma_b^2 = 0)$
Leaf specific area (SLA, *m2/kg*)	0.70**	0.26**	**0.002**
Leaf circularity (Dimensionless)	1.00***	0.71***	**0.001**
Leaf thickness (*mm*)	0.96***	1.80***	**0.001**
Leaf width (*cm*)	0.98***	0.56***	**0.008**
§Animal dispersal (Yes or no)	0.65***	0.28**	**0.054**
Life cycle (Annual or non-annual)	0.00	0.30	0.479
Growth habit (woody or non-woody)	1.08***	0.24**	0.500
Pollination mode (Biotic or abiotic)	0.00	0.08	0.500
Seed mass (*g/seed*)	0.56	0.30	0.373
Leaf dry mass content (LDMC, %)	0.51	0.16	0.500
Stem dry mass content (SDMC, %)	0.00	0.14	0.500
Plant height (*cm*)	0.71**	0.17**	0.500
Leaf length (*cm*)	0.96***	0.32**	0.500
Leaf carbon content (%)	0.65**	0.26**	0.500
Leaf nitrogen content (%)	0.34	0.09	0.334
Wind dispersal (Yes or no)	1.17***	0.46***	0.265
Unassisted dispersal (Yes or no)	0.00	0.15	0.500

531 * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
532

533 **Table 3** Reduction of the phylogenetic variance in community composition caused by the inclusion of

534 functional traits (equation 3).

Trait	σ^2_c with traits	σ^2_c without traits	$100 \times \sigma^2_{c\ (\text{with traits})}/\sigma^2_{c\ (\text{without traits})}$
Leaf width	0.005302	0.006457	17.89
Leaf thickness	0.005921	0.006457	8.30
SLA	0.006024	0.006457	6.71
Leaf circularity	0.006310	0.006457	2.28
Animal dispersal	0.006380	0.006456	1.18
SLA + circularity + thickness + Leaf width + Animal dispersal	0.002804	0.006480	56.73

535

536

537 **Table 4** Variation in the response of species abundances to environmental variables (equation 4).
538 Although 13/20 environmental variables generated variation in species composition among communities,
539 none of these showed phylogenetic signal in which related species responded more similarly to the
540 environmental variable.
541

Environmental variables	*P*-values σ^2_g (no phylogenetic signal)	*P*-values for σ^2_h (phylogenetic signal)
Minimum temperature	< **0.001**	0.500
Precipitation	< **0.001**	0.500
Canopy shade	**0.002**	0.500
Total exchange capacity	**0.002**	0.500
Organic matter	**0.001**	0.500
pH	< **0.001**	0.500
N	< **0.001**	0.500
P	**0.039**	0.500
Mg	**0.030**	0.500
K	**0.007**	0.500
Na	< **0.001**	0.500
Mn	< **0.001**	0.354
Ca	< **0.001**	0.122
Clay	0.110	-
Silt	0.070	-
Sand	0.117	-
Fe	0.500	-
S	0.458	-
Zn	0.500	-
Al	0.500	-

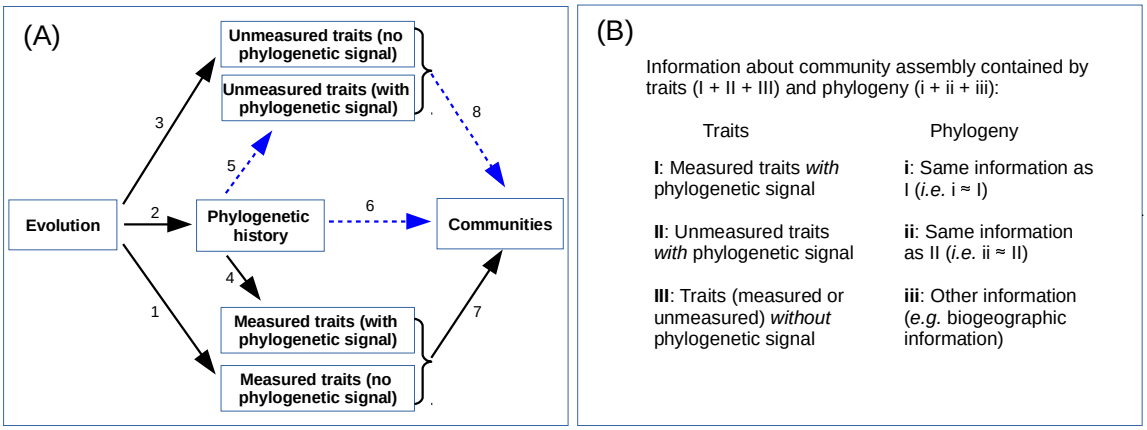
542

Figures:

Figure 1 Schematic diagram of the conceptual framework of the study. (A) Evolution is the ultimate source of all trait values, although only some traits have phylogenetic signal that reflects phylogenetic history (arrows 2, 4 and 5). Other traits do not (arrows 1 and 3), possibly because these traits evolve rapidly or experience convergent evolution. Community composition is determined by unmeasured and measured traits, and also by additional processes that could generate phylogenetic signal, such as biogeographical patterns in the distribution of species. Phylogenetic patterns in community composition can be generated from measured and unmeasured traits with phylogenetic signal (arrows 7 and 8), and by other phylogenetic processes (arrow 6). The question we address is how much of the phylogenetic signal in community composition can be explained by measured functional traits, and whether after accounting for these traits there is residual phylogenetic signal that could have been generated by unmeasured traits or other phylogenetic processes. (B) Traits and phylogeny contain overlapping and complementary information about how communities are assembled. Here, we focus on estimating the proportion of this overlapping information that the phylogeny contains (i.e., the magnitude of *i* relative to *i* + *ii* + *iii*). Note that we do not try to explain the proportion of overlapping information that functional traits contain (i.e., the magnitude of *I* relative to *I* + *II* + *III*) due to our inability to estimate the amount of information provided by unmeasured traits and hence estimate (*I* + *II* + *III*).

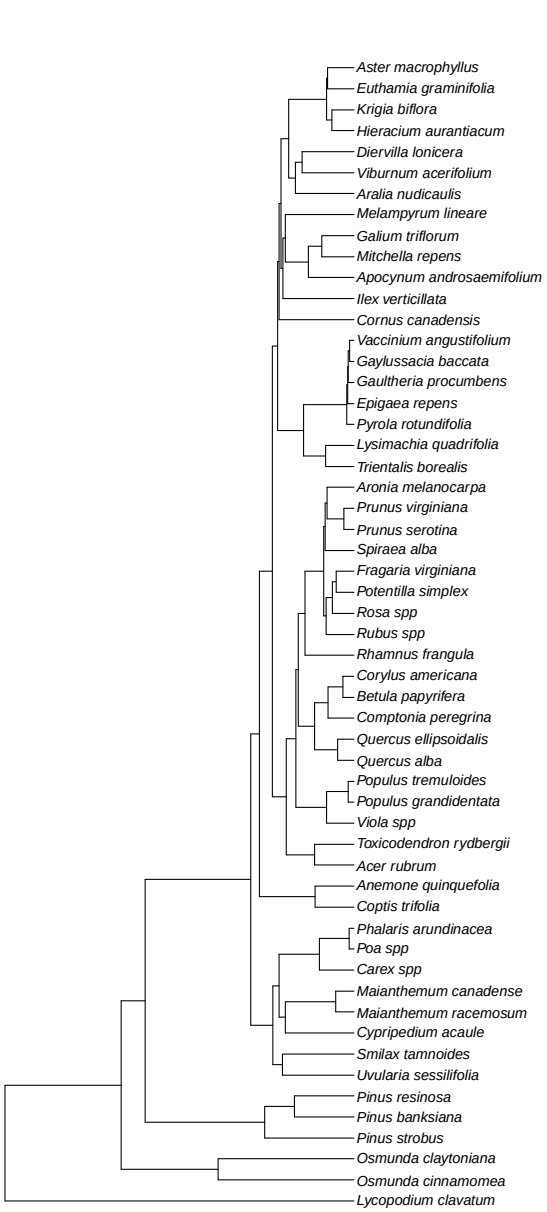
Figure 2: Phylogeny and relative abundance of the 55 common plant species found in the pine barrens of central Wisconsin in 2012. The area of dots is proportional to abundances within each site.

566



567

568 Figure 1



569

570 Figure 2

571

Appendix

In the Appendix we give Tables S1-S4 that correspond to Tables 1-4 in the main text, but using a PGLMM for presence/absence data. The equations used for the PGLMM are the same as equations 1-4, but for binomial data; for example, the PGLMM corresponding to equation 1 is

$$\Pr(Y_i = 1) = \text{logit}^{-1}(\alpha + a_{\text{spp}[i]} + b_{\text{spp}[i]} + c_i + d_{\text{site}[i]}),$$

with other terms identical.

Table S1 Estimated variance of random effects within the phylogenetic generalized linear mixed model used to detect phylogenetic patterns comparable to equation 1, where phylogenetic attraction and phylogenetic repulsion are estimated by σ^2_c.

PGLMM	σ^2_a	σ^2_b	σ^2_c	σ^2_d	$p(\sigma^2_c = 0)$
Phylogenetic attraction: $c \sim \text{Gaussian}(\mathbf{0}, \text{kron}(\mathbf{I}_m, \sigma^2_c\mathbf{\Sigma}_{\text{spp}}))$	2.84	0	0.0452	0.01	<0.001
Phylogenetic repulsion: $c \sim \text{Gaussian}(\mathbf{0}, \text{kron}(\mathbf{I}_m, \sigma^2_c (\mathbf{\Sigma}_{\text{spp}})^{-1})$	3.10	0	0.0011	0.19	0.5
Non-nested model: c removed	2.83	0	-	0.18	-

Table S2 Phylogenetic signal and site variation for each functional trait. P-values for the null hypothesis $\sigma_b^2 = 0$ (equation 2) implying no difference among sites in the effects of trait values on presence/absence are given in the column labeled $p(\sigma_b^2 = 0)$. Functional traits with strong phylogenetic signal and $p(\sigma_b^2 = 0) < 0.1$ are considered to be important in explaining phylogenetic patterns.

Trait	Pagel's λ	*K*	$p(\sigma_b^2 = 0)$
Leaf specific area (SLA, *m2/kg*)	0.70**	0.26**	**0.005**
Leaf circularity (Dimensionless)	1.00***	0.71***	**0.005**
Leaf thickness (*mm*)	0.96***	1.80***	**0.000**
Leaf width (*cm*)	0.98***	0.56***	**0.002**
Animal dispersal (Yes or no)	0.65***	0.28**	**0.002**
Wind dispersal (Yes or no)	1.17***	0.46***	**0.020**
Life cycle (Annual or non-annual)	0.00	0.30	**0.000**
Growth habit (woody or non-woody)	1.08***	0.24**	0.500
Pollination mode (Biotic or abiotic)	0.00	0.08	0.500
Seed mass (*g/seed*)	0.56	0.30	**0.005**
Leaf dry mass content (LDMC, %)	0.51	0.16	0.500
Stem dry mass content (SDMC, %)	0.00	0.14	0.500
Plant height (*cm*)	0.71**	0.17**	0.500
Leaf length (*cm*)	0.96***	0.32**	0.500
Leaf carbon content (%)	0.65**	0.26**	0.282
Leaf nitrogen content (%)	0.34	0.09	0.500
Unassisted dispersal (Yes or no)	0.00	0.15	0.500

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table S3 Proportion of phylogenetic signal of species composition in communities explained by individual functional trait and multiple functional traits. With selected multiple functional traits, about 61% percent of phylogenetic variation was explained, suggesting that phylogenies can provide additional information about community assembly beyond measured functional traits. See equation 3 in the Methods section for details about models.

Trait	σ_c^2 with trait	σ_c^2 without trait	$100 \times \sigma^2_{c\ (\text{with trait})}/\sigma^2_{c\ (\text{without trait})}$
Leaf width	0.018105	0.041847	56.74
Leaf thickness	0.030925	0.041847	26.10
Leaf circularity	0.035442	0.041854	15.32
SLA	0.036811	0.041828	12.00
Wind dispersal	0.039946	0.041844	4.54
Animal dispersal	0.041534	0.041862	0.78
Leaf width + Leaf thickness + Leaf circularity + SLA + Wind dispersal + Animal dispersal	0.004616	0.041834	88.97

Table S4 There are strong variations in species' relationships between their presence/absence and most environmental variables (*p* value of each environmental variable was presented in the *P*-values for variation column). Four of these variations show phylogenetic signal. For environmental variable that has no strong variation in species' responses, no further test for phylogenetic signal of variation was conducted (thus "-" in the third column). P-values that are less than 0.05 are in bold.

Environmental variables	P-values for variation	P-values for phylogenetic signal of variation
Minimum temperature	**<0.001**	**0.002**
Precipitation	**<0.001**	0.500
Canopy shade	**0.001**	0.500
Total exchange capacity	0.149	-
Organic matter	0.161	-
pH	**0.005**	**<0.001**
N	0.052	-
P	0.343	-
Mg	0.500	-
K	0.206	-
Na	**0.004**	0.500
Mn	**<0.001**	**<0.001**
Ca	**0.012**	**<0.001**
Clay	0.431	-
Silt	0.494	-
Sand	0.500	-
Fe	0.379	-
S	0.500	-
Zn	0.500	-
Al	0.500	-

Text S1 Code to compare *p*-values of null hypothesis $\sigma^2 = 0$ calculated from the $0.5\chi_0^2 + 0.5\chi_1^2$ mixture distribution and parametric bootstrap. The *p*-values based on the mixture Chi-square distribution are conservative (i.e. higher than those from parametric bootstrap).

```
# packages used
library(ape)  # for phylogeny reading
library(plyr)
library(MASS)
library(dplyr, quietly = TRUE)
library(pez)  # for communityPGLMM function
library(parallel)  # for multiple cores parallel computation, not available
# for Windows operation system

# data: vegetation data, phylogeny
load("d_li_data.RData")
# select 20 sites and 20 species of veg data in 1958 as an example
test = veg.aggr.wide.1958[1:20, 1:20]
test1 = filter(veg.aggr.long.1958, sp %in% names(test), site %in% rownames(te
st))

# this function calculates log likelihood of the fitted model on observed
# data, then simulates data based on the fitted model, and fits model on
# simulated data and calculates the log likelihood of the fitted model; then
# calculates the p-value of the log likelihood of the fitted model on
# observed data based all simulated ones (i.e. parametric bootstrap); so we
# can compare the p-value get in this way (parametric bootstrap) with the
# one from the mixture Chi-square distribution.
q1_obs_sim = function(veg.long, phylo = pb.phylo, date = 1958, trans = NULL,
    nsim = 100, ncores = 5) {
    # transformation of freq
    if (!is.null(trans)) {
        if (trans == "log") {
            veg.long$Y <- log(veg.long$freq + 1)
        }

        if (trans == "asin") {
            veg.long <- group_by(veg.long, site) %>% mutate(Y = asin(sqrt((fr
eq + 1)/ifelse(date == 1958, 20 + 2, 50 + 2)))) %>% ungroup() %>%
                as.data.frame()
        }
    }

    veg.long$sp = as.factor(veg.long$sp)
    veg.long$site = as.factor(veg.long$site)
    nspp <- nlevels(veg.long$sp)
    nsite <- nlevels(veg.long$site)

    # Var-cov matrix for phylogeny
    phy <- drop.tip(phylo, tip = phylo$tip.label[!phylo$tip.label %in% levels
(veg.long$sp)])
```

```
    Vphy <- vcv(phy)
    Vphy <- Vphy[order(phy$tip.label), order(phy$tip.label)]
    Vphy <- Vphy/max(Vphy)
    Vphy <- Vphy/det(Vphy)^(1/nspp)
    Vphy.inv = solve(Vphy)

    show(c(nlevels(veg.long$sp), Ntip(phy)))  # should be equal

    # random effect for site
    re.site <- list(1, site = veg.long$site, covar = diag(nsite))
    re.sp <- list(1, sp = veg.long$sp, covar = diag(nspp))
    re.sp.phy <- list(1, sp = veg.long$sp, covar = Vphy)
    # sp is nested within site, to test phylo attraction or repulsion
    re.nested.phy <- list(1, sp = veg.long$sp, covar = Vphy, site = veg.long$
site)
    re.nested.rep <- list(1, sp = veg.long$sp, covar = Vphy.inv, site = veg.l
ong$site)

    z <- communityPGLMM(Y ~ 1, data = veg.long, family = "gaussian", sp = veg
.long$sp, site = veg.long$site, random.effects = list(re.sp, re.sp.phy, re.si
te, re.nested.phy), REML = F, verbose = F, s2.init = 0.1)
    show(z$ss)
    z0 <- communityPGLMM(Y ~ 1, data = veg.long, family = "gaussian", sp = ve
g.long$sp, site = veg.long$site, random.effects = list(re.sp, re.sp.phy, re.s
ite),
        REML = F, verbose = F, s2.init = 0.1)
    z.rep <- communityPGLMM(Y ~ 1, data = veg.long, family = "gaussian", sp =
 veg.long$sp, site = veg.long$site, random.effects = list(re.sp, re.sp.phy, r
e.site, re.nested.rep), REML = F, verbose = F, s2.init = 0.1)
    show(z.rep$ss)

    # observed ouput, p-values are get from Chisq approx.
    output_obs = data.frame(LRT_attract = (z$logLik - z0$logLik), p_attract =
 pchisq(2 * (z$logLik - z0$logLik), df = 1, lower.tail = F)/2, LRT_repulse =
(z.rep$logLik - z0$logLik), p_repulse = pchisq(2 * (z.rep$logLik - z0$logLik)
, df = 1, lower.tail = F)/2, obs_sim = "obs")


    # the fitting model z0: log(y_i + 1) = alpha + a_spp[i] +
    # b_spp.phy[i] + c_site[i] + err[i]

    alpha = z0$B  # intercept, overall mean of all sp
    alpha.se = z0$B.se  # SE
    LRT_sim = mclapply(1:nsim, function(x) {
        # multi-cores
        set.seed(x)
        # z0$ss: random effects' SD for the cov matrix \sigma^2 * V, in order
: [1]
        # sp with no phylo; [2] sp with Vphy; [3] site random effect
        a_spp = rnorm(nspp, 0, z0$ss[1])  # simulate a_spp
```

```
        # simulate b_spp.phy
        b_spp.phy = MASS::mvrnorm(1, mu = rep(0, nspp), Sigma = z0$ss[2] * Vphy)
        mu_spp = alpha + a_spp + b_spp.phy  # mean freq of sp
        c_site = rnorm(nsite, 0, z0$ss[3])  # site random
        mu_spp_site = rep(mu_spp, nsite) + rep(c_site, each = nspp)  # each sp at each site
        y_i = rnorm(nspp * nsite, mean = mu_spp_site, sd = alpha.se)  # include SE of intercept
        y_i_count = ceiling(exp(y_i) - 1)  # exp transf and round to positive interge
        test1_sim = data.frame(sp = names(mu_spp_site), site = rep(1:nsite,
            each = nspp), Y = y_i, freq = y_i_count)

        test1_sim$sp = as.factor(test1_sim$sp)
        test1_sim$site = as.factor(test1_sim$site)

        # refit models on simulated data random effect for site
        re.site.sim <- list(1, site = test1_sim$site, covar = diag(nsite))
        re.sp.sim <- list(1, sp = test1_sim$sp, covar = diag(nspp))
        re.sp.phy.sim <- list(1, sp = test1_sim$sp, covar = Vphy)
        # sp is nested within site
        re.nested.phy.sim <- list(1, sp = test1_sim$sp, covar = Vphy, site = test1_sim$site)
        re.nested.rep.sim <- list(1, sp = test1_sim$sp, covar = Vphy.inv, site = test1_sim$site)

        z_sim <- communityPGLMM(Y ~ 1, data = test1_sim, family = "gaussian",

            sp = test1_sim$sp, site = test1_sim$site, random.effects = list(re.sp.sim, re.sp.phy.sim, re.site.sim, re.nested.phy.sim), REML = F, verbose = F, s2.init = 0.1)
        # show(z_sim$ss)
        z0_sim <- communityPGLMM(Y ~ 1, data = test1_sim, family = "gaussian", sp = test1_sim$sp, site = test1_sim$site, random.effects = list(re.sp.sim,
                re.sp.phy.sim, re.site.sim), REML = F, verbose = F, s2.init = 0.1)

        # show(z0_sim$ss)
        z.rep_sim <- communityPGLMM(Y ~ 1, data = test1_sim, family = "gaussian", sp = test1_sim$sp, site = test1_sim$site, random.effects = list(re.sp.sim, re.sp.phy.sim, re.site.sim, re.nested.rep.sim), REML = F, verbose = F, s2.init = 0.1)
        # show(z.rep_sim$ss)

        # log lik of refitted models on simulated data
        data.frame(LRT_attract = (z_sim$logLik - z0_sim$logLik), LRT_repulse = (z.rep_sim$logLik - z0_sim$logLik))
    }, mc.cores = ncores)
```

```
        # output results
    list(output_obs, ldply(LRT_sim))
}

qqq = q1_obs_sim(test1, trans = "log", nsim = 1000, ncores = 6)
saveRDS(qqq, "qqq.rds")

qqq = readRDS("qqq.rds")
qqq[[1]]

##   LRT_attract p_attract  LRT_repulse p_repulse obs_sim
## 1   0.3006013 0.2190598 -1.82719e-05       0.5     obs

head(qqq[[2]])

##   LRT_attract  LRT_repulse
## 1  -0.9611412 -15.68606765
## 2   0.1303866  -0.14712624
## 3  -2.9661583   0.06437319
## 4  -0.2152182  -1.91538503
## 5  -0.4204626   0.04073069
## 6  -1.3125998  -0.40523844

qqq[[2]]$obs_sim = "sim"
q1_sim = rbind(select(qqq[[1]], -p_attract, -p_repulse), qqq[[2]])
1 - (rank(q1_sim$LRT_attract)[1] + 1)/1001  # 0.12088 vs 0.219 from Chisq

## [1] 0.1208791

1 - (rank(q1_sim$LRT_repulse)[1] + 1)/1001  # 0.40959 vs 0.5 from Chisq

## [1] 0.4095904
```